

# Trends Prediction Using Social Diffusion Models

Yaniv Altshuler, Wei Pan, Alex (“Sandy”) Pentland

MIT Media Lab

{yanival,panwei,sandy}@media.mit.edu

The importance of the ability to predict trends in social media has been growing rapidly in the past few years with the growing dominance of social media in our everyday’s life. Whereas many works focus on the detection of anomalies in networks, there exist little theoretical work on the prediction of the likelihood of anomalous network pattern to globally spread and become “trends”. In this work we present an analytic model the social diffusion dynamics of spreading network patterns. Our proposed method is based on information diffusion models, and is capable of predicting future trends based on the analysis of past social interactions between the community’s members. We present an analytic lower bound for the probability that emerging trends would successful spread through the network. We demonstrate our model using two comprehensive social datasets — the *Friends and Family* experiment that was held in MIT for over a year, where the complete activity of 130 users was analyzed, and a financial dataset containing the complete activities of over 1.5 million members of the *eToro* social trading community.

## I. INTRODUCTION

We live in the age of social computing. Social networks are everywhere, exponentially increasing in volume, and changing everything about our lives, the way we do business, and how we understand ourselves and the world around us. The challenges and opportunities residing in the social oriented ecosystem have overtaken the scientific, financial, and popular discourse.

In this paper we study the evolution of trend spreading dynamics in social networks. Where there have been numerous works studying the topic of anomaly detection in networks (social, and others), literature still lacks a theoretic model capable of predicting *how do network anomalies evolve*. When do they spread and develop into global trends, and when they are merely statistical phenomena, local fads that get quickly forgotten? We give an analytically proven lower bound for the spreading probability, capable of detecting “future trends” – spreading behavior patterns that are likely to become prominent trends in the social network.

We demonstrate our model using social networks from two different domains. The first is the *Friends and Family* experiment [1], held in MIT for over a year, where the complete activity of 130 users was analyzed, including data concerning their calls, SMS, MMS, GPS location, accelerometer, web activity, social media activities, and more. The second dataset contains the complete financial transactions of the *eToro community* members – the world’s largest “social trading” platform, allowing users to trade in currency, commodities and indices by selectively copying trading activities of prominent traders.

The rest of the paper is organized as follows : Section II discusses related works. The information diffusion model is presented in Section III and its applicability is demonstrated in Section IV, and concluding remarks are given in Section V.

## II. RELATED WORK

**Diffusion Optimization.** Analyzing the spreading of information has long been the central focus in the study of social networks for the last decade [2] [3]. Researchers have explored both the offline networks structure by asking and incentivizing users to forward real mails and E-mails [4], and online networks by collecting and analyzing data from various sources such as *Twitter* feeds [5].

The dramatic effect of the network topology on the dynamics of information diffusion in communities was demonstrated in works such as [6] [7]. One of the main challenges associated with modeling of behavioral dynamics in social communities stems from the fact that it often involves stochastic generative processes. While simulations on realizations from these models can help us explore the properties of networks [8], a theoretical analysis is much more appealing and robust. In this work we present results are based on a pure theoretical analysis.

The identity and composition of an initial “seed group” in trends analysis has also been the topic of much research. Kempe et al. applied theoretical analysis on the seeds selection problem [9] based on two simple adoption models: *Linear Threshold Model* and *Independent Cascade Model*. Recently, Zaman et al. developed a method to trace rumors back in the topological spreading path to identify sources in a social network [10], and suggest such method can be

used to locate influencers in a network. Some scholars express their doubts and concerns for the influencer-driven viral marketing approach, suggesting that “everyone is an influencer” [11], and companies “should not rely on it” [12]. They argue that the content of the message is also important in determining its spreads, and likely the adoption model we were using is not a good representation for the reality.

Our work, on the other hand, focuses on predicting emerging trends given a current snapshot of the network and adoption status, rather than finding the most influential nodes. We provide a lower bound for the probability that an emerging trend would spread throughout the network, based on the the analysis of the diffusion process outreach, which is largely missing in current literature.

**Adoption Model.** A fundamental building block in trends prediction that is not yet entirely clear to scholars is the adoption model, modeling individuals’ behavior based on the social signals they are exposed to. Centola has shown both theoretical and empirically that a complex contagion model is indeed more precise for diffusion [13, 14]. Different adoption models can dramatically alter the model outcome [15]. In fact, a recent work on studying mobile application diffusions using mobile phones demonstrated that in real world the diffusion process is a far more complicated phenomenon, and a more realistic model was proposed in [16]. Our results also incorporate this realistic diffusion model.

**Trends Prediction and Our Proposed Model.** In this work we study the following question : Given a snapshot of a social network with some behavior occurrences (i.e. an emerging trend), what is the probability that these occurrences (seeds) will result in a viral diffusion and a wide-spread trend (or alternatively, dissolve into oblivious). Though this is similar to the initial seed selection problem [9], we believe that the key factor to succeed in a viral marketing campaign optimization is a better analytical model for the diffusion process itself.

The main innovation of our model is the fact that it is based on a fully analytical framework with a scale-free network model. Therefore, we manage to overcome the dependence on simulations for diffusion processes that characterizes most of the works in this field [6, 17]. We are able to do so by decomposing the diffusion process to the transitive random walk of “exposure agents” and the local adoption model based on [16]. While there are some works that analyze scale-free network [18] most of them come short to providing accurate results, due to the fact that they calculate the expected values of the global behaviors dynamics. Due to strong “network effect” however, many real world networks display much less coherent patterns, involving local fluctuations and high variance in observed parameters, rendering such methods highly inaccurate and sometimes impractical. Our analysis on the other hand tackle this problem by modeling the diffusion process on scale-free networks in a way which takes into account such interferences, and can bound their overall effect on the network.

### III. TREND PREDICTION IN SOCIAL NETWORKS

One of the main difficulty of trends-prediction stems from the fact that the first spreading phase of “soon to be global trends” demonstrate significant similarity to other types of anomalous network patterns. In other words, given several observed anomalies in a social network, it is very hard to predict which of them would result in a wide-spread trend and which will quickly dissolve into oblivious.

We model the community, or social network, as a graph  $G$ , that is comprised of  $V$  (the community’s members) and  $E$  (social links among them). We use  $n$  to denote the size of the network, namely  $|V|$ . In this network, we are interested in predicting the future behavior of some observed anomalous pattern  $a$ . Notice that  $a$  can refer to a growing use of some new web service such as *Groupon*, or alternatively a behavior such as associating oneself with the “99% movement”.

Notice that “exposures” to trends are transitive. Namely, an “exposing” user generates “exposure agents” which can be transmitted on the network’s social links to “exposed users”, which can in turn transmit them onwards to their friends, and so on. We therefore model trend’s exposure interactions as movements of random walking agents in a network. Every user that was exposed to a trend  $a$  generates  $\beta$  such agents, on average.

We assume that our network is (or can be approximated by) a scale free network  $G(n, c, \gamma)$ , namely, a network of  $n$  users where the probability that user  $v$  has  $d$  neighbors follows a power law :

$$P(d) \sim c \cdot d^{-\gamma}$$

We also define the following properties of the network :

**Definition 1.** Let  $V_a(t)$  denote the group of network members that at time  $t$  advocate the behavior associated with the potential trend  $a$ .

**Definition 2.** Let us denote by  $\beta > 0$  the average “diffusion factor” of a trend  $a$ . Namely, the average number of friends a user who have been exposed to the trend will be talking about the trend with (or exposing the trend in other ways).

**Definition 3.** Let  $P_\Delta$  be defined as the probability that two arbitrary members of the network vertices have degrees ratio of  $\Delta$  or higher :

$$P_\Delta \triangleq \text{Prob}[deg(u) > \Delta \cdot deg(v)]$$

**Definition 4.** We denote by  $\sigma_-$  the “low temporal resistance” of the network :

$$\forall t, \Delta_t, \quad \sigma_- \triangleq \max \left\{ 1 \leq \Delta \mid 1 - e^{-\Delta \cdot \frac{\beta \Delta_t \cdot |V_a(t)|}{n}} \cdot (1 - P_\Delta) \right\}$$

**Definition 5.** Let  $P_{Local-Adopt}(a, v, t, \Delta_t)$  denote the probability that at time  $t + \Delta_t$  the user  $v$  had adopted trend  $a$  (for some values of  $t$  and  $\Delta_t$ ). This probability may be different for each user, and may depend on properties such as the network’s topology, past interactions between members, etc.

**Definition 6.** Let  $P_{Local}$  denote that expected value of the local adoption probability throughout the network :

$$P_{Local} = \mathbb{E}_{u \in V} [P_{Local-Adopt}(a, u, t, \Delta_t)]$$

**Definition 7.** Let us denote by  $P_{Trend}(\Delta_t, \frac{|V_a(t)|}{n}, \varepsilon)$  the probability that at time  $t + \Delta_t$  the group of network members that advocate the trend  $a$  has at least  $\varepsilon \cdot n$  members (namely, that  $|V_a(t + \Delta_t)| \geq \varepsilon \cdot n$ ).

We assume that the seed group of members that advocate a trend at time  $t$  is randomly placed in the network. Under this assumption we can now present the main result of this work : the lower bound over the prevalence of an emerging trend. Note that we use  $P_{Local-Adopt}$  as a modular function in order to allow future validation in other environments. The explicit result is given in Theorem 2.

**Theorem 1.** For any value of  $\Delta_t, |V_a(t)|, n, \varepsilon$ , the probability that at time  $t + \Delta_t$  at least  $\varepsilon$  portion of the network’s users would advocate trend  $a$  is lower bounded as follows :

$$P_{Trend} \left( \Delta_t, \frac{|V_a(t)|}{n}, \varepsilon \right) \geq P_{Local}^{\varepsilon \cdot n} \cdot \left( 1 - \Phi \left( \frac{\sqrt{n} \cdot (\varepsilon - \tilde{P}_-)}{\sqrt{\tilde{P}_- (1 - \tilde{P}_-)}} \right) \right)$$

where :

$$\tilde{P}_- = e^{-\left( \frac{\Delta_t \cdot \sigma_-}{2} - \rho_{opt-} + \frac{\rho_{opt-}^2}{2\Delta_t \cdot \sigma_-} \right)}$$

and where :

$$\rho_{opt-} \triangleq \underset{\rho}{\text{argmin}} \left( P_{Local}^{\varepsilon \cdot n} \cdot P_{Trend} \left( \Delta_t, \frac{|V_a(t)|}{n}, \varepsilon \right) \right)$$

and provided that :

$$\rho_{opt-} < \Delta_t \cdot \sigma_-$$

and as  $\sigma_-$  depends on  $P_\Delta$ , using the following bound :

$$\forall v, u \in V, \quad P_\Delta \leq \frac{c^2 \cdot \Delta^{1-\gamma}}{2\gamma^2 - 3\gamma + 1}$$

*Proof.* See Appendix for the complete proof of the Theorem □

Recent studies examined the way influence is being conveyed through social links. In [16] the probability of network users to install applications, after being exposed to the applications installed by the friends, was tested. This behavior was shown to be best modeled as follows, for some user  $v$  :

$$P_{Local-Adopt}(a, v, t, \Delta_t) = 1 - e^{-(s_v + p_a(v))} \quad (1)$$

Exact definitions and methods of obtaining the values of  $s_v$  and  $w_{v,u}$  can be found in [16]. The intuition of these network properties is as follows :

For every member  $v \in V$ ,  $s_v \geq 0$  captures the individual susceptibility of this member, regardless of the specific behavior (or trend) in question.  $p_a(v)$  denotes the *network potential* for the user  $v$  with respect to the trend  $a$ , and is defined as the sum of network agnostic “*social weights*” of the user  $v$  with the friends exposing him with the trend  $a$ .

Notice also that both properties are trend-agnostic. However, while  $s_v$  is evaluated once for each user and is network agnostic,  $p_a(v)$  contributes network specific information and can also be used by us to decide the identity of the network’s members that we would target in our initial campaign.

Using Theorem 1 we can now construct a lower bound for the success probability of a campaign, regardless of the specific value of  $\rho$  :

**Theorem 2.** *For every  $\Delta_t$ ,  $|V_a(t)|$ ,  $n$ ,  $\varepsilon$ , the probability that at time  $t + \Delta_t$  at least  $\varepsilon$  portion of the network’s users advocate the trend  $a$  is :*

$$P_{Trend} \left( \Delta_t, \frac{|V_a(t)|}{n}, \varepsilon \right) \geq e^{-\varepsilon \cdot n \cdot \xi_G \cdot \xi_N^{\rho_{opt-}}} \cdot \left( 1 - \Phi \left( \sqrt{n} \cdot \frac{\varepsilon - \tilde{P}_-}{\sqrt{\tilde{P}_- (1 - \tilde{P}_-)}} \right) \right)$$

where :

$$\tilde{P}_- = e^{-\left(\frac{\Delta_t \cdot \sigma_-}{2} - \rho_{opt-} + \frac{\rho_{opt-}^2}{2\Delta_t \cdot \sigma_-}\right)}$$

and where :

$$\rho_{opt-} \triangleq \underset{\rho}{\operatorname{argmin}} \left( e^{-\varepsilon \cdot n \cdot \xi_G \cdot \xi_N^{\rho}} \cdot P_{Trend} \left( \Delta_t, \frac{|V_a(t)|}{n}, \varepsilon \right) \right)$$

and provided that :

$$\rho_{opt-} < \Delta_t \cdot \sigma_-$$

and where  $\xi_G$  denotes the network’s adoption factor and  $\xi_N$  denotes the network’s influence factor :

$$\xi_G = e^{-\frac{1}{n} \sum_{v \in V} s_v} \quad , \quad \xi_N = e^{-\frac{1}{n} \sum_{e(v,u) \in E} \left( \frac{w_{u,v}}{|N_v|} + \frac{w_{v,u}}{|N_u|} \right)}$$

*Proof.* See complete proof in the Appendix. □

#### IV. EXPERIMENTAL RESULTS

We have validated our model using two comprehensive datasets, the *Friends and Family* dataset that studied the casual and social aspects of a small community of students and their friends in Cambridge, and the *eToro* dataset — the entire financial transactions of over 1.5M users of a “social trading;; community.

The datasets were analyzed using the model given in [16], based on which we have experimentally calculated the values of  $\beta$ ,  $\xi_G$ ,  $\xi_N$  and  $\sigma_-$ .

Figures 1 and 2 demonstrate the probabilistic lower bound for trend emergence, as a function of the overall penetration of the trend at the end of the time period, under the assumption that the emerging trend was observed in 5% of the population. In other words, for any given “magnitude” of trends, what is the probability that network phenomena that are being advocated by 5% of the network, would spread to this magnitude. Notice that although a longer spreading time slightly improve the penetration probability, the “maximal outreach” of the trend (the maximal rate of global adoption, with sufficient probability) is dominated by the topology of the network, and the local adoption features.

#### V. CONCLUSIONS AND FUTURE WORK

In this work we have discussed the problem of trends prediction, that is — observing anomalous network patterns and predicting which of them would become a prominent trend, spreading successfully throughout the network. We have analyzed this problem using information diffusion techniques, and have presented a lower bound for the probability of a pattern to become a global trend in the network, for any desired level of spreading. In order to model the local interaction between members, we have used the results from [16] that studied the local social influence dynamics between members of social networks.

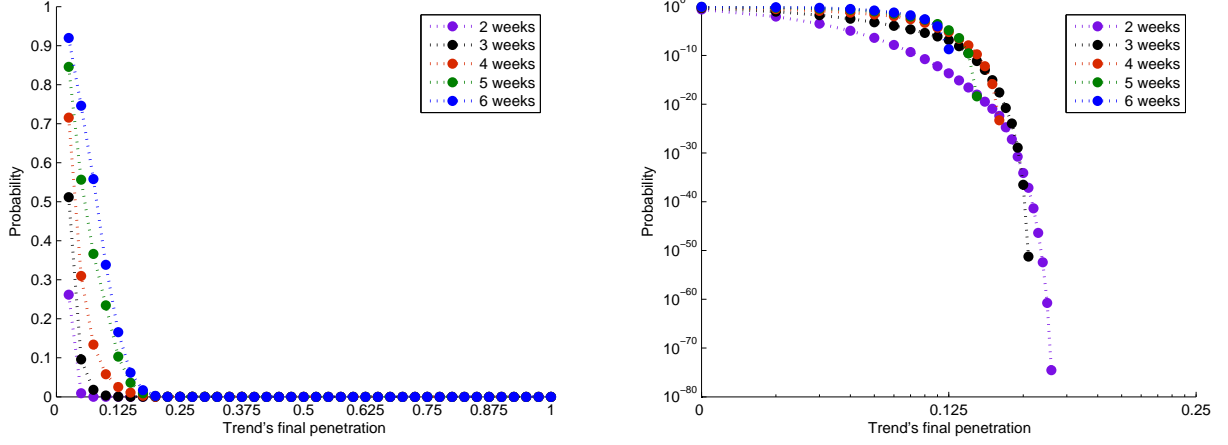


FIG. 1: Trends spreading potential in the *eToro* network, for various penetration rates. Initial seed group is defined as 5% of the population. Each curve represents a different time period, from 2 weeks to 6 weeks.

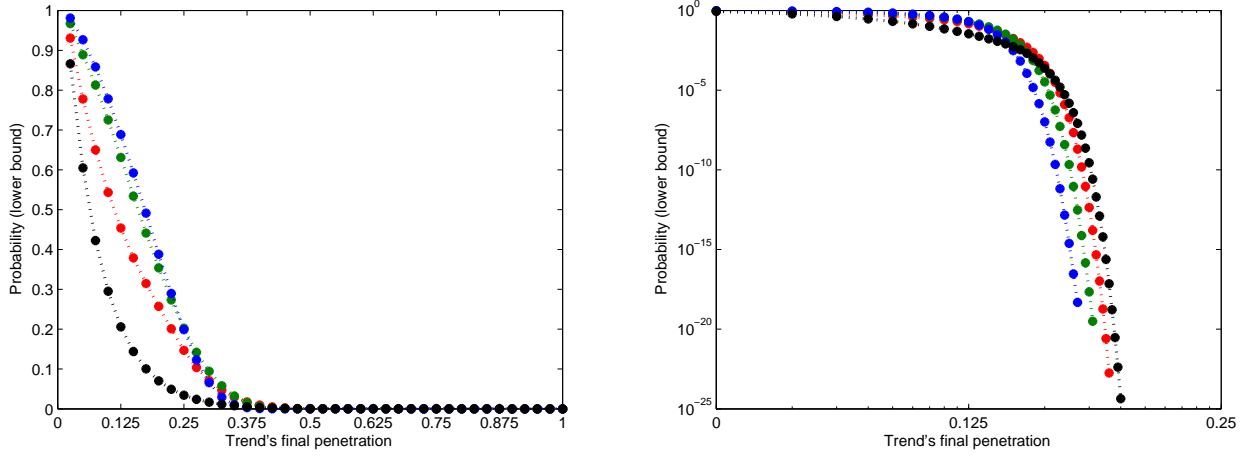


FIG. 2: Trends spreading potential in the *Friends and Family* network, for various penetration rates. Initial seed group is defined as 5% of the population. Each curve represents a different time period, from 2 weeks to 5 weeks.

Though our work provides a comprehensive theoretical framework to understand trends diffusion in social networks, there are still a few challenges ahead. For example, we wish to extend our model to other network models such as Erdos-Renyi random networks, as well as Small World networks. This is essential as more evidences are suggesting that some communities involve complex structures that cannot be easily approximated using a simplistic scale-free model [19].

In addition, our results can be used in order to provide answers to other questions, such as what is the optimal group of members that should be used as a “seed group” in order to maximize the effects of marketing campaigns. Another example might be finding changes in the topology of the social network that would influence the information diffusion progress in a desired way (either to encourage or surpass certain emerging trends).

In order to achieve these goals we are planning a large-scale field test with a leading online social platform, that would give us access to collect more empirical supporting evidences, as well as conducting an active experiment in which we would try to predict trends in real time.

Finally, we are interested in comparing the prediction obtained from our model with the actual semantics of the

trends, to better understand the connection between the trends semantics and the diffusion process they undergo.

- 
- [1] N. Aharony, W. Pan, C. Ip, I. Khayal, A. Pentland, Social fmri: Investigating and shaping social mechanisms in the real world, *Pervasive and Mobile Computing*.
  - [2] B. Huberman, D. Romero, F. Wu, Social networks that matter: Twitter under the microscope, *First Monday* 14 (1) (2009) 8.
  - [3] J. Leskovec, L. Backstrom, J. Kleinberg, Meme-tracking and the dynamics of the news cycle, in: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Citeseer*, 2009, pp. 497–506.
  - [4] P. Dodds, R. Muhamad, D. Watts, An experimental study of search in global social networks, *Science* 301 (5634) (2003) 827.
  - [5] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media?, in: *Proceedings of the 19th international conference on World wide web, ACM*, 2010, pp. 591–600.
  - [6] H. Choi, S. Kim, J. Lee, Role of network structure and network effects in diffusion of innovations, *Industrial Marketing Management* 39 (1) (2010) 170–177.
  - [7] V. Nicosia, F. Bagnoli, V. Latora, Impact of network structure on a model of diffusion and competitive interaction, *EPL (Europhysics Letters)* 94 (2011) 68009.
  - [8] C. Herrero, Ising model in scale-free networks: A monte carlo simulation, *Physical Review E* 69 (6) (2004) 067109.
  - [9] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM*, 2003, pp. 137–146.
  - [10] D. Shah, T. Zaman, Rumors in a network: Who's the culprit?, *Arxiv preprint arXiv:0909.4370*.
  - [11] E. Bakshy, J. Hofman, W. Mason, D. Watts, Everyone's an influencer: quantifying influence on twitter, in: *Proceedings of the fourth ACM international conference on Web search and data mining, ACM*, 2011, pp. 65–74.
  - [12] D. Watts, J. Peretti, H. B. School, Viral marketing for the real world.
  - [13] D. Centola, M. Macy, Complex contagions and the weakness of long ties, *American Journal of Sociology* 113 (3) (2007) 702.
  - [14] D. Centola, The spread of behavior in an online social network experiment, *science* 329 (5996) (2010) 1194.
  - [15] P. Dodds, D. Watts, Universal behavior in a generalized model of contagion, *Physical Review Letters* 92 (21) (2004) 218701.
  - [16] W. Pan, N. Aharony, A. Pentland, Composite social network for predicting mobile apps installation, in: *AAAI*, 2011.
  - [17] S. Banerjee, S. Mallik, I. Bose, Reaction diffusion processes on random and scale-free networks, *Arxiv preprint cond-mat/0404640*.
  - [18] S. Meloni, A. Arenas, Y. Moreno, Traffic-driven epidemic spreading in finite-size scale-free networks, *Proceedings of the National Academy of Sciences* 106 (40) (2009) 16897.
  - [19] J. Leskovec, J. Kleinberg, C. Faloutsos, Graphs over time: densification laws, shrinking diameters and possible explanations, in: *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, ACM*, 2005, pp. 177–187.

# Appendix

## Appendix A: Proof of Theorem 1

In order to prove Theorem 1 we shall require the following definitions :

**Definition 8.** Let  $\mathcal{N}_{v,a}(t)$  denote the number of friends of user  $v$  that at time  $t$  are exposing  $v$  to the trend  $a$  (namely, the number of friends of  $v$  that at time  $t$  have been exposed to the trend  $a$  and are conveying this information to  $v$ ).

Note that “exposing” a neighbor to a trend does not necessarily mean advocating the same trend.

**Definition 9.** Let us denote by  $P_{\rho-Trend}(\Delta_t, \rho, \frac{|V_a(t)|}{n}, \varepsilon)$  the probability that at time  $t + \Delta_t$  at least  $\varepsilon \cdot n$  members of the network have been exposed to the trend  $a$  by at least  $\rho$  of their friends.

In addition, we define  $\rho_{opt-}$  that is used in the Theorem :

**Definition 10.**

$$\rho_{opt-} \triangleq \underset{\rho}{\operatorname{argmin}} \left( P_{Local}^{\varepsilon \cdot n} \cdot P_{Trend} \left( \Delta_t, \frac{|V_a(t)|}{n}, \varepsilon \right) \right)$$

We later see that the expression for  $P_{Trend}$  would refer to  $\rho$ . Using  $\rho_{opt-}$  we would later be able to omit this dependance.

**Theorem 1.**

For any value of  $\Delta_t$ ,  $|V_a(t)|$ ,  $n$ ,  $\varepsilon$ , the probability that at time  $t + \Delta_t$  at least  $\varepsilon$  portion of the network’s users would advocate trend  $a$  is lower bounded as follows :

$$P_{Trend} \left( \Delta_t, \frac{|V_a(t)|}{n}, \varepsilon \right) \geq P_{Local}^{\varepsilon \cdot n} \cdot \left( 1 - \Phi \left( \frac{\sqrt{n} \cdot (\varepsilon - \tilde{P}_-)}{\sqrt{\tilde{P}_- (1 - \tilde{P}_-)}} \right) \right)$$

where :

$$\tilde{P}_- = e^{-\left( \frac{\Delta_t \cdot \sigma_-}{2} - \rho_{opt-} + \frac{\rho_{opt-}^2}{2\Delta_t \cdot \sigma_-} \right)}$$

and provided that :

$$\rho_{opt-} < \Delta_t \cdot \sigma_-$$

*Proof.* We first assess the number of “agents” residing in adjacent vertices from some vertex  $v$  at any given time :

**Lemma 1.** Let  $v \in V$  be an arbitrary user of the network  $G$ . Then :

$$P \left[ E_v[\mathcal{N}_{v,a}(t + \Delta_t) - \mathcal{N}_{v,a}(t)] = \eta \cdot \beta^{\Delta_t} \cdot \frac{|V_a(t)|}{n} \right] = c \cdot \eta^{-\gamma}$$

*Proof.* We assume that the movement of the agents in the network are random[? ]. Hence :

$$\forall u \in V(G), \quad E[\text{number of agents residing on } u] = \frac{\# \text{ of agents}}{n}$$

At time  $t$  there are  $|V_a(t)|$  members of the network that advocate trend  $a$ . Those members generate on average  $\beta$  “agents” that are sent along the social links to their friends, creating chains of length  $\Delta_t$ , and a total of  $\beta^{\Delta_t} \cdot |V_a(t)|$  active agents. Incorporating this with the distribution of the degrees, the rest is implied.  $\square$

The following Lemma produces the probability that two arbitrarily selected vertices would have degrees which differ in more than a certain threshold :

**Lemma 2.**

$$\forall v, u \in V, \quad P_{\Delta} \leq \frac{c^2 \cdot \Delta^{1-\gamma}}{2\gamma^2 - 3\gamma + 1}$$

*Proof.* By definition :

$$P_{\Delta} = \text{Prob}[deg(u) \geq \Delta \cdot deg(v)] \leq$$

$$\int_1^{\infty} \int_{\Delta}^{\infty} (\text{Prob}[deg(v) = j] \cdot \text{Prob}[deg(u) = m \cdot j] dm) dj$$

As the network is scale free, we can write :

$$P_{\Delta} = \text{Prob}[deg(u) \geq \Delta \cdot deg(v)] \leq$$

$$\int_1^{\infty} \int_{\Delta}^{\infty} (c \cdot j^{-\gamma} \cdot c \cdot (m \cdot j)^{-\gamma} dm) dj \leq$$

$$\int_1^{\infty} \left( \frac{c^2 \cdot j^{-2\gamma} \cdot m^{-\gamma+1}}{1-\gamma} \Big|_{m=\Delta}^{\infty} \right) dj \leq$$

$$- \int_1^{\infty} \frac{c^2 \cdot j^{-2\gamma} \cdot \Delta^{-\gamma+1}}{1-\gamma} dj \leq - \frac{c^2 \cdot j^{-2\gamma+1} \cdot \Delta^{1-\gamma}}{(1-2\gamma)(1-\gamma)} \Big|_{j=1}^{\infty} \leq$$

$$\frac{c^2 \cdot \Delta^{1-\gamma}}{(2\gamma-1)(\gamma-1)} \leq \frac{c^2 \cdot \Delta^{1-\gamma}}{2\gamma^2-3\gamma+1}$$

□

**Lemma 3.** For any member  $v \in V$  at time  $t + \Delta_t$ , the probability that  $v$  will be exposed at the next time-step to the trend  $a$  is at least  $\sigma_-$ .

*Proof.* The probability that an agent located on a vertex  $u$  such that  $(u, v) \in E$  will move to  $v$  at the next time-step is  $\frac{1}{deg(u)}$ . Therefore, remembering that  $v$  has  $\mathcal{N}_{v,a}(t)$  agents that resides in neighboring vertices in time  $t$ , the probability that  $v$  will *not* be exposed to the trend at the next time-step is :

$$\left( 1 - \frac{1}{deg(u)} \right)^{\mathcal{N}_{v,a}(t)} \tag{A1}$$

Using the well known inequality  $(1-x) < e^{-x}$  for  $x < 1$ , and taking into account Lemma 1, we can bound Equation A1 from above by :

$$\left( 1 - \frac{1}{deg(u)} \right)^{\mathcal{N}_{v,a}(t)} \leq e^{-\frac{\mathcal{N}_{v,a}(t)}{deg(u)}} \leq e^{-\frac{deg(v)}{deg(u)} \cdot \frac{\beta \Delta_t \cdot |V_a(t)|}{n}} \tag{A2}$$

Using Lemma 2 we can further simplify Equation A2 as follows :

$$\text{Prob} \left[ e^{-\frac{deg(v)}{deg(u)} \cdot \frac{\beta \Delta_t \cdot |V_a(t)|}{n}} \leq e^{-\Delta \cdot \frac{\beta \Delta_t \cdot |V_a(t)|}{n}} \right] \geq (1 - P_{\Delta}) \tag{A3}$$

Therefore, the probability that a user will be exposed to the trend in the next time step is at least :

$$1 - e^{-\Delta \cdot \frac{\beta \Delta_t \cdot |V_a(t)|}{n}} \cdot (1 - P_{\Delta})$$

which equals  $\sigma_-$ .

□

We can now proceed to the calculation of  $P_{\rho-Trend}$ .



**Lemma 4.** *The probability that at time  $t + \Delta_t$  at least  $\varepsilon \cdot n$  members of the network have been exposed to a trend  $a$  by at least  $\rho$  of their friends is lower bounded as follows :*

$$P_{\rho-Trend}(\Delta_t, \rho, \frac{|V_a(t)|}{n}, \varepsilon) \geq 1 - \Phi \left( \sqrt{n} \cdot \frac{\varepsilon - \tilde{P}_-}{\sqrt{\tilde{P}_-(1 - \tilde{P}_-)}} \right)$$

where :

$$\tilde{P}_- = e^{-(\frac{\Delta_t \cdot \sigma_-}{2} - \rho + \frac{\rho^2}{2\Delta_t \cdot \sigma_-})}$$

and where  $\Phi(x)$  is the cumulative normal distribution function, defined as :

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt$$

and also provided that :

$$\rho < \Delta_t \cdot \sigma_-$$

*Proof.* Using Lemma 3 we have a lower bound for the probability that a user  $v$  will be exposed to some trend  $a$  by an agent originated by one of the group of users that advocate the trend  $a$  at time  $t$ . This is in fact a *Bernoulli* trial with success probability of  $\sigma_-$ .

Denoting  $X_v(t)$  the number of times user  $v$  is being exposed to the trend  $a$  after  $t$  steps, we shall now use the negative variance *Chernoff* bound :

$$P[X_v(t) < (1 - \delta)t \cdot \sigma_-] < e^{-\delta^2 \frac{t \cdot \sigma_-}{2}}$$

Once selecting  $\delta = 1 - \frac{\rho}{t \cdot \sigma_-}$  and for the entire lifespan of the trend (namely, for  $t = \Delta_T$ ) we obtain the probability that a single (specific) member will be exposed to the trend  $a$  at least  $\rho$  times. For this, we shall first define :

$$\tilde{P}_- \triangleq P_{\rho-Trend}(\Delta_T, \rho, \frac{|V_a(t)|}{n}, n^{-1})$$

which by definition implies :

$$\begin{aligned} \tilde{P}_- &= P[X_v(\Delta_T) < \rho] < e^{-(1 - \frac{\rho}{\Delta_T \cdot \sigma_-})^2 \frac{\Delta_T \cdot \sigma_-}{2}} < \\ &e^{-(1 - 2\frac{\rho}{\Delta_T \cdot \sigma_-} + \frac{\rho^2}{\Delta_T \cdot \sigma_-^2}) \frac{\Delta_T \cdot \sigma_-}{2}} < e^{-(\frac{\Delta_T \cdot \sigma_-}{2} - \rho + \frac{\rho^2}{2\Delta_T \cdot \sigma_-})} \end{aligned}$$

As the *Chernoff* bound requires that  $\delta > 0$  we should make sure that :

$$\rho < \Delta_T \cdot \sigma_-$$

As we want to bound the probability that at least  $\varepsilon \cdot n$  of the network members are exposed to the trend at least  $\rho$  times, we shall use the above as a success probability of yet a second *Bernoulli* trial. As  $n$  is assumed to be large, the number of exposed members can be approximated using *Normal* distribution :

$$P_{\rho-Trend} \left( \Delta_T, \rho, \frac{|V_a(t)|}{n}, \varepsilon \right) \geq 1 - \Phi \left( \frac{\varepsilon \cdot n - n \cdot \tilde{P}_-}{\sqrt{n \cdot \tilde{P}_-(1 - \tilde{P}_-)}} \right)$$

and the rest is implied. □

Whereas Lemma 4 provides an estimation concerning the global outreach of trends in terms of exposure, it does not take into account the probability that users that are exposed to the trend by  $\rho$  of their friends, will actually adopt it. In order to do this, we need to incorporate  $P_{Local-Adopt}$  into Lemma 4, producing a combined bound for the global adoption of the trend.

**Proposition 1.** For any  $\Delta_t$ ,  $|V_a(t)|$ ,  $n$ ,  $\varepsilon$ , the probability that at time  $t + \Delta_t$  at least  $\varepsilon$  portion of the network's users advocate the legitimate trend  $a$  is :

$$P_{Trend,\rho} \left( \Delta_t, \rho, \frac{|V_a(t)|}{n}, \varepsilon \right) = P_{Local}^{\varepsilon \cdot n} \cdot P_{\rho-Trend} \left( \Delta_t, \rho, \frac{|V_a(t)|}{n}, \varepsilon \right)$$

Notice that  $\rho$  appears in the expression of Proposition 1 for mathematical reasons, and has no actual meaning. We omit the dependency of the expression on  $\rho$ , by finding the optimal value of  $\rho$  for every set of values of  $\varepsilon$ ,  $|V_a(t)|$  and  $\Delta_t$ , by assigning :

$$P_{Trend} \left( \Delta_t, \frac{|V_a(t)|}{n}, \varepsilon \right) = P_{Trend,\rho} \left( \Delta_t, \rho_{opt-}, \frac{|V_a(t)|}{n}, \varepsilon \right)$$

And the rest is implied.  $\square$

## Appendix B: Proof of Theorem 2

Let us remind once again the local influence model that was shown in [16] to best approximate the behavior diffusion in real world social networks :

$$P_{Local-Adopt}(a, v, t, \Delta_t) = 1 - e^{-(s_v + p_a(v))} \quad (B1)$$

We recall that  $s_v \geq 0$  captures the individual susceptibility of this member, and that  $p_a(v)$  denotes the *network potential* for the user  $v$  with respect to the trend  $a$ , and is defined as the sum of network agnostic “*social weights*” of the user  $v$  with the friends exposing him with the trend  $a$  :

$$p_a(v) = \sum_{u \in \mathcal{N}_{v,a}} w_{v,u}$$

(where  $\mathcal{N}_{v,a}$  is the overall group of users exposing  $v$  to the trend  $a$ ).

Using Theorem 1 we can now construct a lower bound for the success probability of a campaign, regardless of the specific value of  $\rho$  :

**Theorem 2.** For every  $\Delta_t$ ,  $|V_a(t)|$ ,  $n$ ,  $\varepsilon$ , the probability that at time  $t + \Delta_t$  at least  $\varepsilon$  portion of the network's users advocate the legitimate trend  $a$  is :

$$P_{Trend} \left( \Delta_t, \frac{|V_a(t)|}{n}, \varepsilon \right) \geq e^{-\varepsilon \cdot n \cdot \xi_G \cdot \xi_N^{\rho_{opt-}}} \cdot \left( 1 - \Phi \left( \sqrt{n} \cdot \frac{\varepsilon - \tilde{P}_-}{\sqrt{\tilde{P}_- (1 - \tilde{P}_-)}} \right) \right)$$

where :

$$\tilde{P}_- = e^{-\left(\frac{\Delta_t \cdot \sigma_-}{2} - \rho_{opt-} + \frac{\rho_{opt-}^2}{2\Delta_t \cdot \sigma_-}\right)}$$

and where :

$$\rho_{opt-} \triangleq \underset{\rho}{\operatorname{argmin}} \left( e^{-\varepsilon \cdot n \cdot \xi_G \cdot \xi_N^{\rho}} \cdot P_{Trend} \left( \Delta_t, \frac{|V_a(t)|}{n}, \varepsilon \right) \right)$$

and provided that :

$$\rho_{opt-} < \Delta_t \cdot \sigma_-$$

and where  $\xi_G$  denotes the network's *adoption factor* and  $\xi_N$  denotes the network's *influence factor* :

$$\xi_G = e^{-\frac{1}{n} \sum_{v \in V} s_v}$$

$$\xi_N = e^{-\frac{1}{n} \sum_{e(v,u) \in E} \left( \frac{w_{u,v}}{|\mathcal{N}_{v,a}|} + \frac{w_{v,u}}{|\mathcal{N}_{u,a}|} \right)}$$

*Proof.* From Equation B1 we have :

$$P_{Local-Adopt}(a, v, t, \Delta_t) = 1 - e^{-(s_v + p_a(v))}$$

The expected value of the local adoption probability is therefore :

$$E_{u \in V}[P_{Local-Adopt}(u)] = \frac{1}{n} \sum_{v \in V} 1 - e^{-(s_v + \frac{\rho}{|\mathcal{N}_{v,a}|} \sum_{u \in \mathcal{N}_{v,a}} w_{v,u})}$$

(where  $\mathcal{N}_{v,a}$  is the groups of user  $v$ 's friends).

Using the inequality  $(1 - x) < e^{-x}$  for  $x < 1$ , we see that :

$$P_{Local-Adopt}^{\varepsilon \cdot n} = \left( \frac{1}{n} \sum_{v \in V} 1 - e^{-(s_v + \frac{\rho}{|\mathcal{N}_{v,a}|} \sum_{u \in \mathcal{N}_{v,a}} w_{v,u})} \right)^{\varepsilon \cdot n} = \quad (B2)$$

$$\left( 1 - \frac{1}{n} \sum_{v \in V} e^{-(s_v + \frac{\rho}{|\mathcal{N}_{v,a}|} \sum_{u \in \mathcal{N}_{v,a}} w_{v,u})} \right)^{\varepsilon \cdot n} <$$

$$e^{-\varepsilon \cdot \sum_{v \in V} e^{-(s_v + \frac{\rho}{|\mathcal{N}_{v,a}|} \sum_{u \in \mathcal{N}_{v,a}} w_{v,u})}}$$

Using the fact that an arithmetic mean is always greater than a geometric mean, Equation B2 can be written as follows :

$$P_{Local-Adopt}^{\varepsilon \cdot n} < e^{-\varepsilon \cdot \sum_{v \in V} e^{-(s_v + \frac{\rho}{|\mathcal{N}_{v,a}|} \sum_{u \in \mathcal{N}_{v,a}} w_{v,u})}} < \quad (B3)$$

$$e^{-\varepsilon \cdot n \cdot \sqrt[n]{\prod_{v \in V} e^{-(s_v + \frac{\rho}{|\mathcal{N}_{v,a}|} \sum_{u \in \mathcal{N}_{v,a}} w_{v,u})}}} <$$

$$e^{-\varepsilon \cdot n \cdot e^{-\frac{1}{n} \sum_{v \in V} (s_v + \frac{\rho}{|\mathcal{N}_{v,a}|} \sum_{u \in \mathcal{N}_{v,a}} w_{v,u})}}} <$$

$$e^{-\varepsilon \cdot n \cdot e^{-\frac{1}{n} \sum_{v \in V} s_v} \cdot e^{-\frac{\rho}{n} \sum_{e(v,u) \in E} (\frac{w_{u,v}}{|\mathcal{N}_{v,a}|} + \frac{w_{v,u}}{|\mathcal{N}_{u,a}|})}}}$$

Integrating Equation B3 with Proposition 1 and Lemma 4 the rest is implied.  $\square$